Keep Querying and Tag on: Collaborative Folksonomy Using Model-Based Recommendation

Angelina de C.A. Ziesemer and João Batista S. de Oliveira

Centro de Pesquisa em Computação Aplicada
Faculdade de Informática — PUCRS
Porto Alegre - RS, Brazil
angelina.ziesemer@acad.pucrs.br,
oliveira@pucrs.br

Abstract. Tags are terms commonly used in collaborative media systems like Flickr, Youtube and Picasa to classify a subject, image, video, music or any related content. Despite its popularity, tagging is a repetitive task and that may affect the quality and reuse of tags in collaborative systems. In this paper we use a model-based tag recommendation approach to perform an experiment and analyze the vocabulary homogeneity of queries (tags provided by users), the recommended tags and their reuse. Results show that the use of recommendation improves the quality and reuse of tags. Furthermore, based on users attribution behavior, we conclude with a proposal for personalized tag recommendation.

Keywords: collaborative filtering, folksonomy, recommendation.

1 Introduction

Having a large amount of information distributed online, the categorization of content became impossible for system administrators. Web 2.0 environments allow users the possibility to categorize items through tags using folksonomy (folk + taxonomy), very popular in social media like Flick, Instagram and YouTube because in photos and videos there is no textual information to be found by search engines.

Despite the advantages of tags, tagging is a repetitive, tedious work and that may affect the quantity of tags and their quality. Sigurbjornsson et al. [1] made a substantial contribution to understanding the long tail on tag distribution: they analyzed how users assign tags, mostly hints to where/who/what and when the photo was taken. In addition, according to Kennedy [2] only 50% of tags provided by users are truly related to the resources. However, collaborative tagging [3] is a powerful tool on social media networks and could be improved by recommender systems [4].

In this paper we present the results from an experiment with queries (tags provided by users during an item classification) and recommended tags (suggested by a recommeder engine) to analyze the reuse and homogeneity of them

P. Antunes et al. (Eds.): CRIWG 2013, LNCS 8224, pp. 10–17, 2013.

when there is a recommender system involved. We implemented a model-based collaborative filtering (CF) [5] approach to recommend tags and compute the utility of them by probability measures using machine learning techniques. The results show that tag recommendation can improve tag reuse and homogeneity. In the next section we present the recommendation model used in the experiment and its improvements.

2 Reviewing and Improving the Recommendation Model

The tag recommender [6] used in this experiment describes each post P_i in a social tag system as a triple $P_i =< u_i, r_i, T_i >$ where $T_i = \{t_1, t_2...t_n\}$ is a set of tags attributed to resource r_i posted by user u_i. A tag t typed by user is treated like a query for similar recommendations based on its co-occurrence in $P(t) = \{P_i | t \in T_i\}$. To develop the recommendation it is necessary to obtain the k-tags with largest co-occurrence from $P(t)$. The function

$$exist(t, T) = \begin{cases} 1, t \in T \\ 0, t \notin T \end{cases} \tag{1}$$

will signal the existence of t in T and is used to rank the co-occurring tags t_j by $ranking(t, t_j) = \sum_{P_i \in P(t)} exist(t_j, T_i)$. That will produce the preliminary ranking of tags to compute the next three measures to improve the recommendation of a tag t_j.

Co-occurrence: To use the ranking of co-occurring tags and to take a normalized value for each tag, we compute the number of items that have both t and t_j by

$$coo(t, t_j) = \frac{ranking(t, t_j)}{|P(t)|} \tag{2}$$

The $coo(t, t_j)$ value for each t_j ranges from 0 to 1.

Relevance: The relevance measure tries to take from the top of the ranking those tags that do not represent the community vocabulary.

$$rel(t, t_j) = \frac{|users(t) \bigcap users(t_j)|}{ranking(t, t_j)} \tag{3}$$

For example, if a user posts lots of photos from a trip to Paris and the same set of tags is used for all of them, $< Paris, France, Mary, Aaron >$, these personal names will appear with a high level of co-occurrence in the ranking. Computing tag relevance will help us sort out tags that occur many times but are attributed only by a user or few users and its value will be low when this behavior occurs.

Popularity: The popularity measure is the number of users using tags t and t_j divided by the number of users that have t and it measures how popular t_j is to users which have t in their resources.

$$pop(t, t_j) = \frac{|users(t) \bigcap users(t_j)|}{|users(t)|} \tag{4}$$

After computing these three measures, the final ranking of recommended tags is computed for all tags in the list of co-occurring tags by the geometric mean:

$$mean(t, t_j) = \sqrt[3]{coo(t, t_j) * rel(t, t_j) * pop(t, t_j)} \tag{5}$$

Most tag recommendation approaches do not take into account that tags could have ambiguous meaning. *AutoTag*[7] uses information retrieval measures to estimate the similarity between weblog posts and then weigh each associated tag based on its frequency to recommend a list of tags for new content. Also, Sigurbjornsson et al. [1] proposed four approaches to address the problem of tag recommendation, also using tag co-occurrence, but it uses only one tag at a time to recommend others.

To filter a better set of tags from recommendation, our approach was improved to use a set $S = \{t_1, t_2, t_3...\}$ of tags to obtain refined results for recommendation. Thus, it will search the k-largest co-occurrence tags for $P(S)$ using S as query:

$$exist(S, T) = \begin{cases} 1, S \subseteq T \\ 0, S \nsubseteq T \end{cases} \tag{6}$$

This will signal the existence of S in the set T for each item r in the data set and rank the co-occurring tags by the function $ranking(S, t_j)$, compute the measures ($coo(S, t_j)$, $rel(S, t_j)$, $pop(S, t_j)$) and perform the geometric mean as presented.

Figure 1 shows how the algorithm uses more than one tag selected by the user: suppose that a user is categorizing an item, provides a tag used as query NY (a) and accepts the tags “statueofliberty”, “statue”, “nyc”, “trip” and “newyork” (b) recommended by the regular approach (using NY to get recommendation). To get more specific tags based on the context of the item, it is possible to use those tags that are in level b (already assigned to the item as in (c)) to obtain refined tags (e) using the set S of tags (d) that where in level b. For example, the combination between “statueofliberty” and “newyork” in level c will return refined results like “manhattan”, “libertyisland”, “usa”, “statenisland” and “newyorkcity” in level e.

The combination of more than one tag to recommend others helps to filter and avoid tags from distinct contexts such as “Paris” from “Paris Hilton” or “Paris” related to “France”. If the user has the chance to indicate the context using sets of tags to get more tags, the recommendation can be more refined. In the next section we present the results of the experiment performed using the model with its improvements.

3 Experiment

To verify if whether use of recommendation improves the reuse and homogeneity of tags in a collaborative system, we performed an experiment using a training data set from Flickr with more than 600.000 tags and 49.120 distinct tags in total. The engine was freely available online for two weeks and 50 participants classified photos from Flickr that do not require effort in recognizing where/who/what is

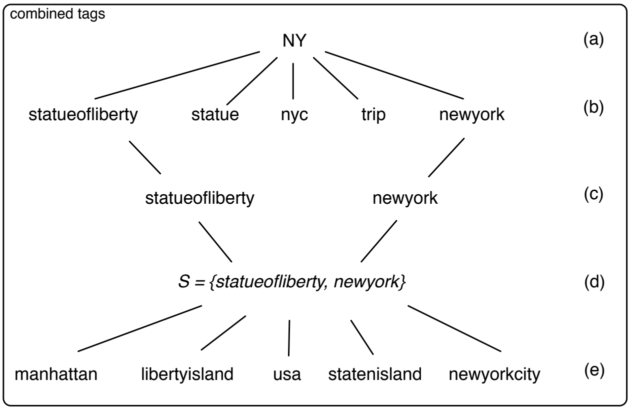

Fig. 1. Improving the model combining tags to get refined results

in the photo. Each tag provided by the users was used as a query to obtain a list of recommended tags (suggested by the engine). We stored all queries used by users and those recommended tags that were accepted by them during the experiment. In the next section, we present the results from the analysis of the data arising from the experiment.

Reuse of Tags. One of the biggest problems in tagging systems is that most of tags are used only one time. The approach presented in this paper tries to provide the best tags for each subject and through its acceptance improve the reuse of tags and the vocabulary homogeneity of items.

Table 1 shows the quantity of queries typed and those tags accepted by users during the experiment. In a total of 891 queries, 182 were distinct and 97 of them were used only one time, showing that 53% of the queries typed by the users were not reused. On the other hand, in the set of 1.235 recommended tags accepted by the users there are 145 distinct tags and 93 of them were reused. Statistically the Z-test of proportion among this sample results in a p-value equal to 0.0484, showing that there is a significant difference in the proportion of reuse of tags accepted by the recommendation comparing to those tags provided by the users and used like queries. In other words, the results show that the reuse of recommended tags is better since 66% of them were reused and only 34% were used a single time.

Table 1. Frequency of tags and queries, the center and spread of dataset resulted from the experiment

Behavior	Total	Distinct	Used Once	Mean	Median	3rd Quartile
Queries	891	182	97 (53%)	4.89	1	3
Tags	1.235	145	52 (34%)	8.51	3	13

Moreover, during the experiment the mean number of queries used was 4.89 and the median was 1.0, showing that at least one half of the queries typed were used only one time. In the other hand the mean of tags accepted by the recommendation was 8.51, almost twice the mean of queries typed by users. Moreover, the median of tags accepted was 3.0. Also, we compute the 3rd quartile of the curve of tags (Table 1) to analyze the spread of the data, with the curve of the queries and recommended tags in Figure 2. The x axis represents each distinct query and recommended tags in the data set and y their frequency.

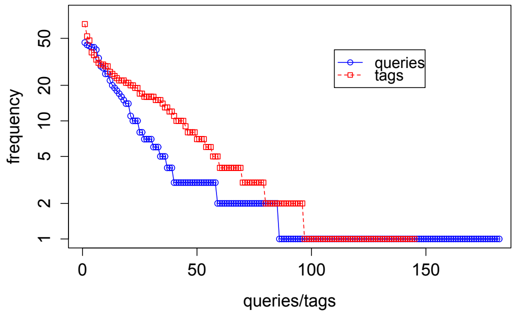

Fig. 2. Long tail of tags provided by the users and recommended tags. Even though the number of recommended tags is higher than the number of queries, the long tail of recommended tags is smaller than queries due to reuse of them.

Comparing the long tail of recommended tags and queries we can better understand the values of the 3rd quartile: for this data the reuse of queries was smaller than the recommended tags. In the next section we report the results from the experiment related to the ranking position of tags during the recommendation and the users behavior related to measures and tagging task.

Tag Ranking Position and Users Behavior. In our algorithm we intend to put on the top of the ranking the most relevant tags through the use of co-occurrence, relevance and popularity measures, thus the ranking position of tags accepted by users is an important point to observe. Figure 3 shows the result of the experiment related to the quantity of recommended tags accepted and their position when they were recommended.

We observe that most of recommended tags accepted are in the first five ranking positions. Moreover, we analyzed the tag recommendation by the tailored precision of tags used in [8], for the first five positions (P@5) compared to the tags presented in the first ten positions (P@10). The precision in five was 0.25 and in ten was 0.18.

However, during the experiment we observed that for one of the images classified by the users, the position of the most accepted tags was not in the five first positions, it was in the eighth place of the ranking. This shows that users choose of tags were not only based on their position on the ranking but in the word itself. Based on these results, the engine frequently brought suitable tags

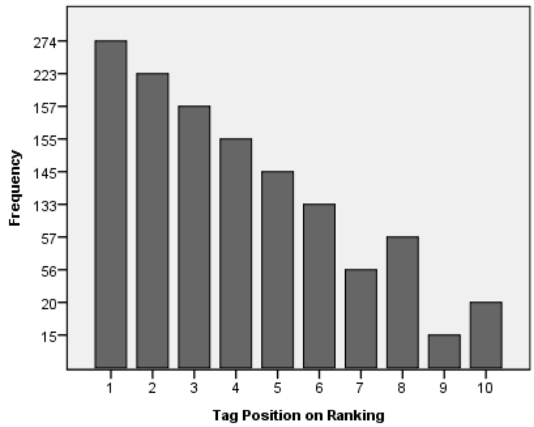

Fig. 3. Frequency and position of recommended tags during the experiment

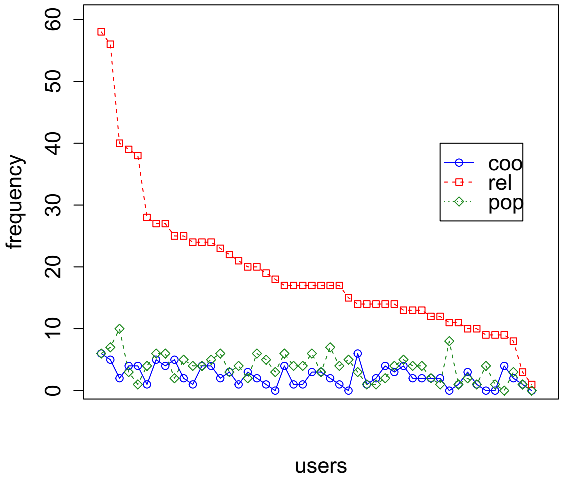

Fig. 4. User's measures behavior by the acceptance of tags

to the top of the ranking. Moreover, the influences of the position of tags it is an important point to be future investigated in comparison with document retrieval approaches.

Also, we verified which tag measures stood out as more important. Each accepted tag has three values (*coo*, *rel*, *pop*) that were used to observe if there is a tag measure that is most significant for each user. Indeed, most tags accepted have a relevance value higher than the co-occurrence and popularity as shown in Figure 4. This observation opens room for personalized measures for recommendation as will be discussed in the next section.

Additionally, we applied a survey to understand what users think about tagging task. The first question was about how frequently users categorize their photos: 48% said that they never use tags to categorize photos. When asked "what are the reasons why people do not tag photos?", most of answers were that "tagging is a hard task" and 99% of them think that recommender tag system helps users to assign tags. Moreover, 41% of users agree that people do not know which tags are good for tagging. This reinforces the importance of collaborative approaches in folksonomy systems and the improvement that can be provided by recommender techniques. In the next section we present a new approach to recommend tags based on the results of this experiment.

4 A Personalized Approach

Based on user's behavior, we may propose a personalized recommendation approach using the measures tied to users' preferences. We still use the three measures for ranking tags, but now their significance is related to previously accepted tags. To personalize the recommendation each profile has a triple $u = < w_c, w_r, w_p >$ where each w is the number of times that the measure was the higher among all measures. For example, to obtain the weight for the relevance we compute w_r by

$$w_r = \sum higher(\text{argmax}\,(coo, rel, pop), rel) \tag{7}$$

where $higher(\text{argmax}\,(x, y, z), x)$ will be 1 if $\text{argmax}\,(x, y, z) = x$ and 0 if it is not. Thereafter, we use the weighted geometric mean where each measure has its corresponding weight to compute the personalized recommendation for each user:

$$mean(t, t_j) = \sqrt[\Sigma w]{coo(t, t_j)^{w_c} * rel(t, t_j)^{w_r} * pop(t, t_j)^{w_p}} \tag{8}$$

Table 2 shows a preliminary result from the personalized approach compared to the recommender approach (non-personalized) used in this paper. To obtain the recommendation we used the weighted measures as $rel > pop > coo$, according to the tags measure users behavior resulted from the experiment. The personalized recommendation approach presents a tag variation that is also suitable for the queries used and it shows that the algorithm can produce distinct and good results using measures based on users attribution history.

Table 2. Comparing the tags recommended using the proposal approaches

Query	**Non-personalized**	**Personalized**
nature	butterfly, bird, wildlife	bird, flower, water
ny	statueofliberty, newyork, nyc	usa, policecar, harbor
beach	sand, ocean, dog	sand, boat, ship
venice	italy, gondola, bridge	gondola, water, street
zoo	polarbear, rhino, penguin	rhino, animal, lion

5 Conclusions and Future Work

Results from this experiment show that collaborative filtering approaches can improve tag reuse and have a positive impact in the vocabulary homogeneity. Further, the proportion of recommended tags attributed was higher than the tags provided by users (queries) and most users in the experiment agree that tagging is a hard task and recommender systems can facilitate it. We intend to perform an experiment using the personalized recommendation against the recommendation shown in this paper. Also, we intend to implement gamification techniques in combination with recommender techniques to improve the user experience and encourage users to use tags more frequently.

References

1. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proceedings of the 17th International Conference on World Wide Web, WWW 2008, pp. 327–336. ACM (2008)
2. Kennedy, L.S., Fu Chang, S., Kozintsev, I.V.: To search or to label?: predicting the performance of search-based automatic image classifiers. In: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pp. 249–258. ACM (2006)
3. Dattolo, A., Ferrara, F., Tasso, C.: The role of tags for recommendation: a survey. In: Proc. of the 3rd International Conference on Human System Interaction - HSI 2010, pp. 548–555. IEEE Press (2010)
4. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowl. and Data Eng. 17, 734–749 (2005)
5. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. In: Adv. in Artif. Intell., vol. 2009, pp. 1–19 (2009)
6. Ziesemer, A., Oliveira, J.B.: Social tag collaborative recommendation: Mining the meaning. In: Proceedings of the Eighth Brazilian Symposium on Collaborative Systems - SBSC 2011, pp. 124–127 (2011)
7. Mishne, G.: Autotag: a collaborative approach to automated tag assignment for weblog posts. In: Proceedings of the 15th International Conference on World Wide Web - WWW 2006, pp. 953–954. ACM (2006)
8. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in social bookmarking systems. AI Commun. 21, 231–247 (2008)